



## Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder

Lescai, F; Als, T D; Li, Q; Nyegaard, M; Andorsdottir, G; Biskopstø, M; Hedemand, A; Fiorentino, A; O'Brien, N; Jarram, A; Liang, J; Grove, J; Pallesen, J; Eickhardt, E; Mattheisen, M; Bolund, L; Demontis, D; Wang, A G; McQuillin, A; Mors, O; Wang, J; Børglum, A D

*Published in:*  
Translational Psychiatry

*DOI:*  
[10.1038/tp.2017.3](https://doi.org/10.1038/tp.2017.3)

*Publication date:*  
2017

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY](#)

*Citation for published version (APA):*  
Lescai, F., Als, T. D., Li, Q., Nyegaard, M., Andorsdottir, G., Biskopstø, M., Hedemand, A., Fiorentino, A., O'Brien, N., Jarram, A., Liang, J., Grove, J., Pallesen, J., Eickhardt, E., Mattheisen, M., Bolund, L., Demontis, D., Wang, A. G., McQuillin, A., ... Børglum, A. D. (2017). Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder. *Translational Psychiatry*, 7, [e1034].  
<https://doi.org/10.1038/tp.2017.3>

## ORIGINAL ARTICLE

## Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder

F Lescai<sup>1,2,3</sup>, TD Als<sup>1,2,3</sup>, Q Li<sup>4</sup>, M Nyegaard<sup>1,2,3</sup>, G Andorsdottir<sup>5</sup>, M Biskopstø<sup>5</sup>, A Hedemand<sup>1,2,3</sup>, A Fiorentino<sup>6</sup>, N O'Brien<sup>6</sup>, A Jarram<sup>6</sup>, J Liang<sup>4</sup>, J Grove<sup>1,2,3,7</sup>, J Pallesen<sup>1,2,3</sup>, E Eickhardt<sup>1,2,3</sup>, M Mattheisen<sup>1,2,3</sup>, L Bolund<sup>1,3,8</sup>, D Demontis<sup>1,2,3</sup>, AG Wang<sup>9</sup>, A McQuillin<sup>4</sup>, O Mors<sup>2,3,8</sup>, J Wang<sup>3,4</sup> and AD Børglum<sup>1,2,3</sup>

Bipolar disorder affects about 1% of the world's population, and its estimated heritability is about 75%. Only few whole genome or whole-exome sequencing studies in bipolar disorder have been reported, and no rare coding variants have yet been robustly identified. The use of isolated populations might help finding variants with a recent origin, more likely to have drifted to higher frequency by chance. Following this approach, we investigated 28 bipolar cases and 214 controls from the Faroe Islands by whole exome sequencing, and the results were followed-up in a British sample of 2025 cases and 1358 controls. Seventeen variants in 16 genes in the single-variant analysis, and 3 genes in the gene-based statistics surpassed exome-wide significance in the discovery phase. The discovery findings were supported by enrichment analysis of common variants from genome-wide association studies (GWAS) data and interrogation of protein–protein interaction networks. The replication in the British sample confirmed the association with *NOS1* (missense variant rs79487279) and *NCL* (gene-based test). A number of variants from the discovery set were not present in the replication sample, including a novel *PITPNM2* missense variant, which is located in a highly significant schizophrenia GWAS locus. Likewise, *PIK3C2A* identified in the gene-based analysis is located in a combined bipolar and schizophrenia GWAS locus. Our results show support both for existing findings in the literature, as well as for new risk genes, and identify rare variants that might provide additional information on the underlying biology of bipolar disorder.

*Translational Psychiatry* (2017) 7, e1034; doi:10.1038/tp.2017.3; published online 14 February 2017

## INTRODUCTION

Bipolar disorder is a disturbance of mood in which patients display episodes of depression, often characterised by low mood, loss of pleasure and energy, and episodes of hypomania or mania, with irritable mood, increased energy and reduced sleep: this condition affects about 1% of the world's population.<sup>1</sup> The estimated heritability of bipolar disorder is about 75%.<sup>2</sup> There is a substantial overlap in genetic aetiology with schizophrenia with several loci influencing susceptibility to both disorders, for example, at *CACNA1C* and *PIK3C2A*.<sup>3–5</sup> Analysis of data from common variants in genome-wide association studies (GWAS) has shown the genetic correlation between the two disorders to be as high as 0.68.<sup>6</sup> Although common variants may explain a large proportion of the variance in liability to bipolar disorder (and other psychiatric disorders), a substantial part of the estimated heritability is still unaccounted for.<sup>6,7</sup> Rare risk variants not effectively assessed by GWAS may explain part of this hidden heritability.

Whole genome or whole-exome sequencing studies have proved successful in identifying rare variants for Mendelian disorders and more recently also for complex disorders.<sup>8</sup> Among psychiatric disorders, in particular exome sequencing studies in autism have successfully identified rare transmitted variants and *de novo* mutations conveying large effect on disease risk.<sup>9,10</sup> In bipolar disorder, only few whole genome or whole-exome

sequencing studies have been reported, most of which investigated large pedigrees, and no rare coding variants have yet been robustly identified.<sup>11</sup>

Owing to the increased genetic drift during founding, followed by population expansion, isolated populations may be particularly useful in identifying rare disease variants that may appear at higher frequencies compared with outbred populations,<sup>12–14</sup> as has been shown previously for several monogenic<sup>15,16</sup> as well as some complex disorders.<sup>17</sup> Variants with a recent origin are thus more likely to have drifted to higher frequency by chance in a smaller isolated population compared with a larger outbred population.<sup>18</sup> Furthermore, isolated populations are relatively homogeneous in genetic background and environmental exposure, and control cohorts might reflect better the composition of the population they are drawn from, as they often represent a larger proportion of that population compared with controls from outbred populations.

The population of the Faroe Islands is an isolated population. It was founded by a small number of individuals in the ninth century and has experienced limited immigration for several centuries.<sup>19</sup> Owing to the extensive founder and drift effects with rare variants drifting to increased frequencies, certain monogenic disorders appear at highly increased frequencies in the population.<sup>20</sup> For instance, in the Faroese population glycogen storage disease III (GSD3, OMIM #232400) is caused by a single

<sup>1</sup>Department of Biomedicine, Aarhus University, Aarhus, Denmark; <sup>2</sup>PSYCH—The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; <sup>3</sup>SEQ—Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; <sup>4</sup>BGI-Shenzhen, Shenzhen, China; <sup>5</sup>Genetic Biobank of the Faroe Islands, Tórshavn, Faroe Islands; <sup>6</sup>Division of Psychiatry, University College London, London, UK; <sup>7</sup>BiRC—Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark; <sup>8</sup>Aarhus University Hospital, Risskov, Aarhus, Denmark and <sup>9</sup>Mental Health Centre Amager, Copenhagen, Denmark. Correspondence: Professor AD Børglum, Department of Biomedicine, Aarhus University, Bartholins Allé 6, Aarhus 8000, Denmark.

E-mail: anders@biomed.au.dk

Received 29 November 2016; accepted 30 November 2016

nonsense mutation that is >250 times more frequent than in outbred populations.<sup>21</sup>

We aimed to identify rare risk variants in bipolar disorder, by investigating whole-exome sequences of 28 individuals with bipolar disorder and 214 controls from the isolated population of the Faroe Islands. The results were followed-up in a British sample of 2025 bipolar disorder cases and 1358 controls, as well as by enrichment analysis of large GWAS data sets and by the analysis of significant physical connectivity among proteins encoded for by genes nominally significant in this study.

## MATERIALS AND METHODS

### Faroese subjects

Patients were included from the Department of Psychiatry, National Hospital in Torshavn, the capital of the Faroe Islands. The patients included were interviewed by trained interviewers using a brief version of Present State Examination. Based on the hospital records and the interview, a clinical description was made of each patient by an experienced psychiatrist. The final diagnosis was made by best-estimate by an experienced psychiatrist (AGW) on the basis of all the material and records, according to ICD-10, Diagnostic Criteria for Research. All patients included had bipolar disorder according to ICD-10 and bipolar disorder type 1 according to DSM-IV. Controls were included by public advertising and assessed as having no psychiatric record and confirming this in a short interview.

Genomic DNA was purified from whole blood samples at the Faroese Genetic Biobank according to standard procedures used in the laboratories.

The study has been approved by the local scientific ethical committee of the Faroe Islands.

### British subjects

The UCL bipolar disorder subjects comprised 2025 subjects suffering from bipolar disorder type I (83%) or bipolar disorder type II. All subjects had been given a National Health Service (NHS) clinical diagnosis of ICD-10 bipolar disorder and then needed to fulfill the criteria for the lifetime version of the Schizophrenia and Affective Disorder Schedule (SADS-L)<sup>22</sup> which provides a research diagnostic criteria (RDC) diagnosis.<sup>23</sup> The UCL control sample comprised in total 1358 subjects. This included 878 subjects with no first-degree family or personal history of psychiatric illness, supplemented with an additional 480 unscreened normal British subjects obtained from the European Collection of Animal Cell Cultures (ECACC). The bipolar subjects and the screened controls underwent ancestral screening to be included only if at least three out of four grandparents were English, Scottish, Welsh or Irish and if the fourth grandparent was non-Jewish European. National Health Service multi-center research ethics approval was obtained. All participants provided signed consent. DNA samples were collected from blood or saliva samples and genomic DNA was purified using standard techniques.

### Sequence processing

**Library and sequencing.** The library preparation was performed according to the manufacturer's instructions, and the exome was captured using Agilent SureSelect version 3 (Agilent Technologies, Santa Clara, CA, USA). The libraries were sequenced on an Illumina HiSeq2500 (Illumina, San Diego, CA, USA).

**Mapping.** The sample reads were aligned to the genome (reference GRCh37) using BWA version 0.7.4 (<http://bio-bwa.sourceforge.net>), converted to BAM format and indexed using SAMtools (version 0.1.18, <https://samtools.github.io>). The samples were re-aligned, marked for duplicates and recalibrated using GATK<sup>24</sup> and Queue (version 2.7-2, <https://software.broadinstitute.org/gatk/>) as pipeline manager.

**Variant calling.** The variants were called using HaplotypeCaller and UnifiedGenotyper, processed with VQSR following the best practices for the version in use, and pre-filtered by 'PASS' at the output of each caller. The calls were merged by including, in order of priority, all 'PASS' HaplotypeCaller variants and all 'PASS' calls unique to UnifiedGenotyper. This implies that (i) in case of overlapping 'PASS' variants, the calls from HaplotypeCaller were included, (ii) in case of overlapping variants filtered

by VQSR in HaplotypeCaller and 'PASS' in UnifiedGenotyper, the UnifiedGenotyper calls were included and (iii) all 'PASS' non-overlapping variants unique to each Caller were included.

**Annotation.** The variants were annotated using the snpEFF (version 3.3 h, <http://snpeff.sourceforge.net>), EPACTS (version 3.3, <http://genome.sph.umich.edu/wiki/EPACTS>) and Variant Effect Predictor (version 75, <http://www.ensembl.org/info/docs/tools/vep/index.html>) tools from ENSEMBL, and the variant type using the GATK VariantAnnotator. On the basis of this, the variant calls were grouped into single-nucleotide polymorphisms (SNPs), insertions, deletions and multiallelic calls. SIFT, Polyphen and Loftee have been used to annotate missense mutation with additional predictions about potentially damaging consequences. The group of multiallelic calls comprise the variant types identified by GATK as 'MULTIALLELIC\_COMPLEX.Other' and 'MULTIALLELIC\_MIXED': the first includes the variants represented by multiple alleles containing insertions or deletions (or a combination hereof) of different sizes, whereas the second includes the variants in which multiple alternative alleles can be a combination of SNPs, insertions and/or deletions.

### Genotyping and validation

Genotyping of the UCL sample and genotype validation on Faroese samples was performed using the Sequenom MassARRAY iPLEX technology<sup>25</sup> or Fluidigm technology.<sup>26</sup> An in-house Perl script was used to process the VCF file with the significant variants and format the polymorphisms according to Sequenom and Fluidigm requirements. The output for Sequenom was used to design the amplicon and extension primers using the Sequenom Assay Design Suite version 1.0 (Sequenom, San Diego, CA, USA) with high multiplexing iPLEX presets. The genotyping was performed according to manufacturer's standard protocols for iPLEX. The variants that failed Sequenom design or genotyping were typed with Fluidigm: the output of our scripts was processed with D3 Assay Design, and the samples were processed according to Fluidigm. In both cases all genotyping results were manually checked to verify the cluster plots.

### Statistical analyses

Pairwise coefficients of Identity By Descent and kinship coefficients ( $k_{ij}$ ) was estimated using the method of moments approach<sup>27</sup> as implemented in the R-package SNPRelate,<sup>28</sup> to identify unknown relationships and confirming known first-degree relationships. These analyses were conducted on a filtered set of single-nucleotide variants (SNVs), filtered by LD-pruning ( $r^2 < 0.002$ ), a missing rate  $< 0.005$  and a minor allele frequency  $> 0.01$ . The cryptic relatedness evident from the results of these analyses (results not shown), was corrected for in the association analysis, by adopting the EMMAX<sup>29</sup> approach as implemented in EPACTS and suggested previously in the literature.<sup>12</sup>

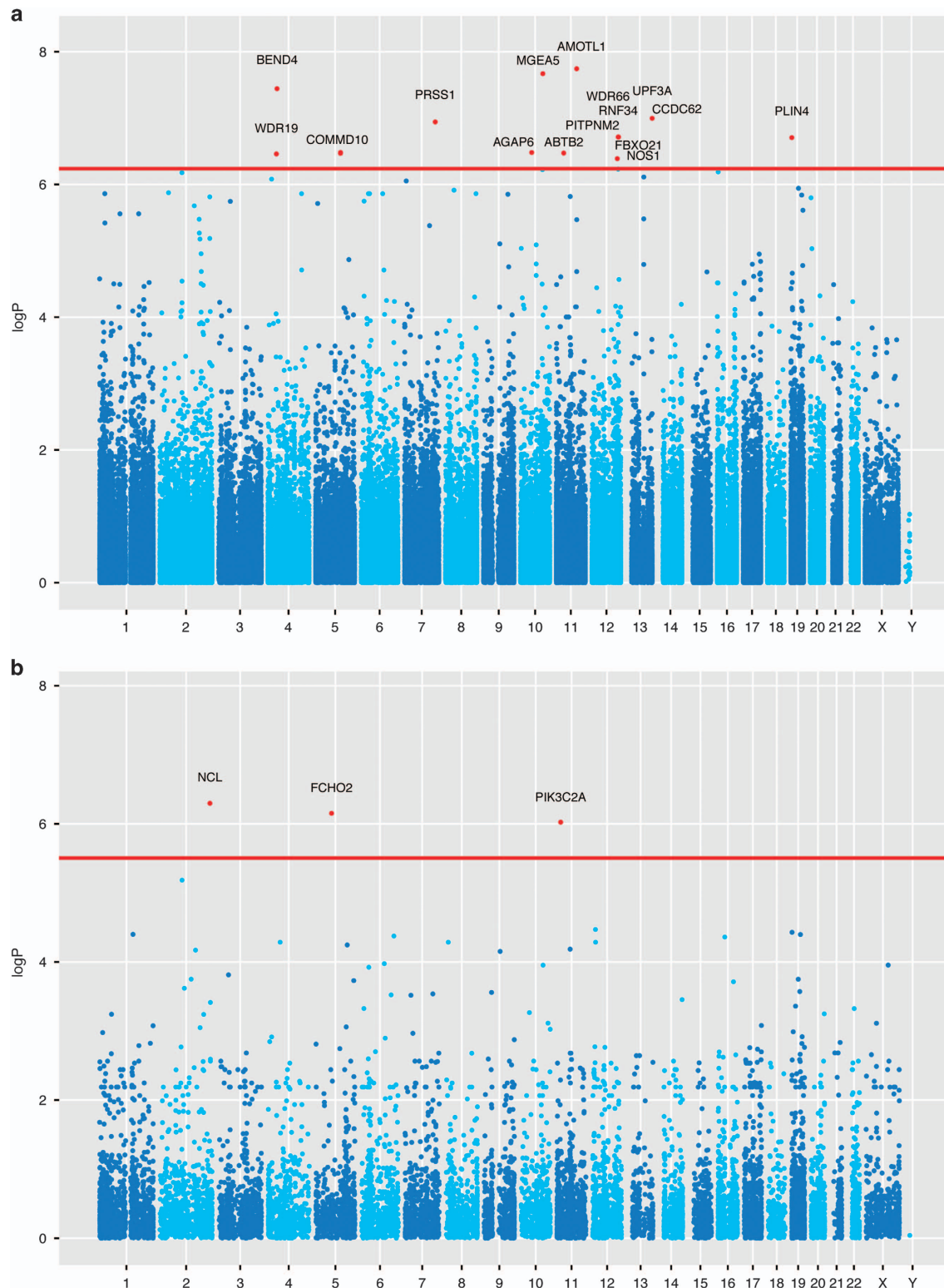
The primary analyses of the Faroese variants data were performed using EPACTS v.2.6. 4 samples have been excluded by EPACTS QC filters (1 case, 3 controls), resulting in a final dataset of 27 cases and 211 controls ready for the analysis. To analyse the single variants according to our hypothesis, we filtered for all called variants either novel or by frequency lower than 0.05 in 1000 Genomes (CEU+GBR) and present in at least 3 individuals in our dataset. Subsequently, the single-variant analysis on the selected subset was performed using the 'q.emmax' statistics,<sup>29</sup> to account for any hidden relatedness of the sample. This method was chosen because of its ability to handle related individuals. The method has been originally developed for quantitative traits, but can be applied to binary traits in the spirit of Armitage trend test giving reliable  $P$ -values but potentially inaccurate effect measures, which we therefore ignore. This is not an exact method but based on asymptotic approximations that may not be accurate when there are small cell counts.

For the gene-burden statistics, the 'Emmax CMC-like' method was employed.<sup>29,30</sup> To collapse the rare variants into gene-loci, the variants were selected by allowing a maximum minor allele frequency (calculated on the entire sample) of 5%, and having the following consequences, as annotated by VEP: transcript\_ablation, splice\_donor\_variant, splice\_acceptor\_variant, stop\_gained, frameshift\_variant, stop\_lost, initiator\_codon\_variant, inframe\_insertion, inframe\_deletion, missense\_variant, transcript\_amplification, splice\_region\_variant, incomplete\_terminal\_codon\_variant.

The statistical analyses of the genotypes from the replication phase for the single-marker analysis were conducted using Fisher's exact test, as implemented in PLINK 1.9.<sup>18</sup>

The replication of the gene-based tests from the British population was performed using the AssotesteR package in R (<http://cran.r-project.org/package=AssotesteR>), and choosing the classic CMC test, as the approach implemented on EPACTS can only be used with a larger genome/exome-wide dataset. To account for multiple testing, empirical *P*-values have been calculated with 1000 permutations, as implemented in the R-package.

For the risk gene enrichment analysis we used MAGMA<sup>31</sup> and default settings, as well as INRICH<sup>32</sup> with default settings, using top-1% GWAS results. INRICH uses a permutation procedure on genomic intervals, whereas MAGMA is based on a multiple regression model. We used publicly available summary statistics from single-marker GWAS<sup>33</sup> considering only variants outside the broad MHC-region (chr6:25M-35M) and



**Figure 1.** The figures report the Manhattan plots of the single-variant analysis with q.emmax (a) and gene-based CMC-like emmax (b) as implemented in EPACTS. The horizontal red line indicates the significant thresholds ( $P$ -value threshold of  $5.78 \times 10^{-7}$  for single-marker tests, and  $3.12 \times 10^{-6}$  for the gene-based tests). Significant variants are annotated with their corresponding gene from ENSEMBL data.



**Table 1.** Selected single variant and gene-burden results.

Gene	Marker	Consequence	P-value Faroese	P-value UK (corrected)	Support in GWAS
<i>NOS1</i>	rs79487279	missense_variant	4.09E-07	0.002 (0.032)	The <i>NOS1</i> locus has a lead signal in schizophrenia with <i>P</i> -value of $1.24 \times 10^{-6}$ (rs2293052) <sup>33</sup>
<i>PITPNM2</i>	12:123489064_C/A	missense_variant	1.93E-07	NA	In highly significant schizophrenia locus with a lead <i>P</i> -value of $2.19 \times 10^{-14}$ (rs2851447) <sup>33</sup>
<i>NCL</i>	Gene test	—	5.04E-07	0.016 (0.029)	—
<i>PIK3C2A</i>	Gene test	—	9.46E-07	NA	<i>PIK3C2A</i> has a significant lead signal with <i>P</i> -value of $6.46 \times 10^{-9}$ in a combined bipolar and schizophrenia GWAS (rs4356203) <sup>39</sup>

The table summarises the most significant results of the study, both from single-variant analysis and gene-based tests. We report only results exome-wide significant in the Faroese population after Bonferroni correction (nominal *P*-values thresholds of  $5.78 \times 10^{-7}$  for the single variants and  $3.12 \times 10^{-6}$  for the gene tests). The *P*-value for the UK sample was calculated with a Fisher exact test on the genotype model with PLINK 1.9, and we report in parenthesis the *P*-value after Bonferroni correction. In the last column we indicate whether our finding is supported by genome-wide association studies (GWAS).

filtered for info score  $\geq 0.8$ . Genes were annotated using Ensemble (GRCh37.p13). Information about the genetic correlation pattern in the data (linkage disequilibrium) was obtained using the 1000 Genomes European panel.<sup>34</sup>

For the protein-protein interaction analysis DAPPLE was used, which looks for significant physical connectivity among proteins encoded by the genes associated in the study. DAPPLE builds interaction networks from proteins encoded by the genes reported in the association study and connected either by direct interactions (that is, when both were significantly associated) or indirect interactions (that is, through proteins not resulting from the association analysis). To identify candidate loci, a score is calculated for each gene by enumerating the number of its connections and comparing this number to the values obtained in permuted networks (50 000 permutations).<sup>35</sup> The *P*-value of this test is the one we refer to in the description of this study.

## RESULTS

Whole exomes from 28 bipolar cases and 214 controls were sequenced at an average depth of  $35\times$ . After mapping the sequences to the GRCh37 version reference of the human genome, the bam files were processed using GATK (see 'Materials and Methods' section), and a total of 259 904 variants were called. These included 230 797 SNVs, 10 029 insertions and 13 198 deletions: among those we annotated 2575 loss-of-function (LoF) variants and 54 967 missense mutations. Overall, 47 800 of the called variants were novel, and were not present in dbSNP.

### Single variant analysis

We designed the study to target rare risk variants (that might have increased in frequency in the isolated Faroese population) and, consequently, we included only variants at frequencies lower than 0.05 (or not present) in the 1000 Genomes data (CEU and GBR samples, release 20110521) (Supplementary Figure 1). In addition, given the small discovery sample size, we limited the analysis to those variants appearing in more than 2 individuals (allele count  $> 2$ ). This filtering strategy produced an analysis-ready dataset of 86 563 variants, corresponding to an experiment/exome-wide significance threshold of  $5.78 \times 10^{-7}$  for single-variant association after Bonferroni correction.

For association analysis we used the mixed model method *q*.emmax,<sup>29</sup> implemented in the software package EPACTS, to adjust for relatedness and population structure within the sample. This analysis resulted in 17 variants in 16 genes surpassing exome-wide significance (Figure 1a). We decided to follow-up all variants with a nominal *P*-value  $< 10^{-6}$  adding up to a total of 24 variants from 18 loci (Supplementary Table 1 and Supplementary Table 2).

As both common and rare risk variants often occur in the same genes/loci,<sup>36,37</sup> we investigated whether the identified top-18 loci

showed risk enrichment for common variants in large bipolar and schizophrenia GWAS data sets,<sup>33,38</sup> using the two methods INRICH and MAGMA.<sup>31,32</sup> Interestingly, we found evidence of enrichment for schizophrenia risk by INRICH (*P*-value=0.031) assessing the top-1% associated SNPs, and by MAGMA (*P*=0.052), supporting the validity of the identified rare variants as a group. This enrichment reflects that several of the identified rare variants are located in loci showing *P*-values in the order of  $10^{-4}$  to  $10^{-6}$  in the GWAS (Table 1). Noteworthy, the *PITPNM2* missense variant is located in a genome-wide significant locus, and it is only described in the ExAC database (<http://exac.broadinstitute.org/>) with a frequency of 0.0006793. Furthermore, the observed enrichment indicates that some of the rare variants identified may influence susceptibility to both bipolar disorder and schizophrenia.

We did not observe any enrichment for bipolar common variant risk, which may be due to a substantially smaller sample size than is the case for the schizophrenia GWAS, providing relatively low power for the enrichment analysis in bipolar disorder.

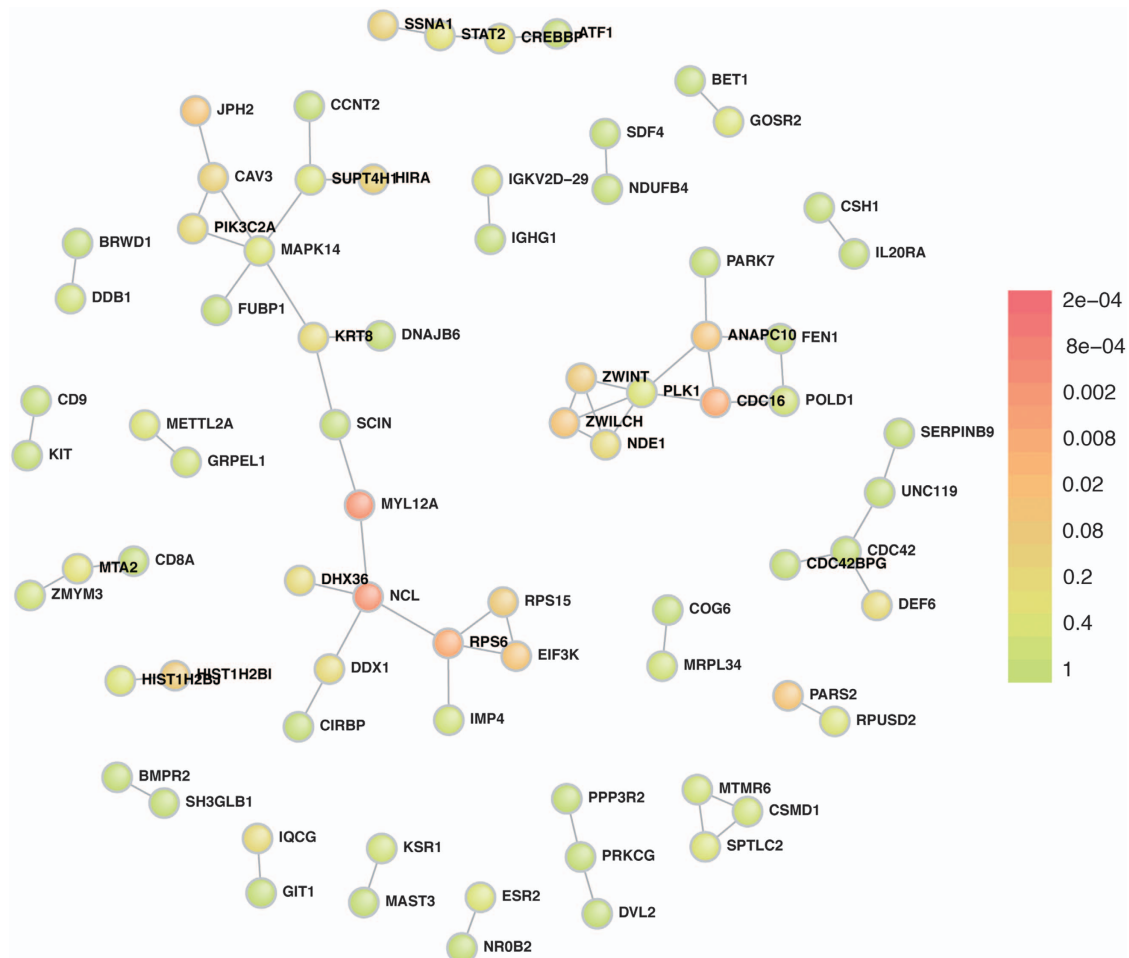
We next followed-up the top-24 rare variants by genotyping a British sample of 2025 bipolar cases and 1358 controls. We chose a British sample as the Faroese population was founded partly by individuals from the British Isles and partly by Scandinavian Vikings,<sup>40</sup> suggesting that risk variants may be shared among the British and Faroese populations.<sup>41</sup> Fifteen variants from 13 loci were successfully genotyped; 9 were either not present in the British cohort or failed genotyping (Supplementary Table 1). The *NOS1* missense variant was the only variant showing significant association withstanding Bonferroni correction (*P*=0.002, *P*<sub>corrected</sub>=0.032; Table 1).

### Gene-wise analysis

To perform collapsing statistics on the rare variants, we collapsed into gene-loci those variants with predicted significant biological effects on the coding regions (see 'Materials and Methods' section) and a maximum minor allele frequency of 5% in the whole dataset. The CMC-like burden test<sup>30</sup> resulted in 419 nominally significant genes at *P*<0.01, and three genes significant exome-wide after Bonferroni correction (Figure 1b) and including more than one rare variant.

In an attempt to replicate the results, we genotyped all variants that contributed to the three significant gene-based tests in the British sample (Supplementary Table 3). However, only in the case of *NCL* were all variants present in the British sample. Thus, a regular replication test could be performed solely for *NCL*, which showed significant association (*P*=0.029, Table 1).

On the nominally significant burden associations (419 genes at *P*-value  $< 0.01$ ), we performed a DAPPLE analysis,<sup>35</sup> investigating



**Figure 2.** The plot shows the most significant connections resulting from a DAPPLE analysis of the 419 nominally significant genes ( $P$ -value  $< 0.01$ ) in the CMC-like emmax burden test in EPACTS of the Faroese sample. The colour code indicates the corrected  $P$ -value range of the DAPPLE analysis (seed scores).

whether the genes implicate a limited set of underlying mechanisms detectable by protein–protein interactions. The analysis supported 16 of the genes from the gene-burden analysis, in terms of protein–protein connectivity (Supplementary Table 4). Remarkably, the most significant gene in this analysis was *NCL*, with a corrected  $P$ -value of 0.002 (Figure 2), which supports the observed association of *NCL* with bipolar disorder, and suggests that more members of its interaction network might be implicated in disease susceptibility.

## DISCUSSION

Targeting rare risk variants we sought to take advantage of using an isolated population, in which some of the variants that are very rare in outbred populations have been found highly increased in frequency, including mutations for rare Mendelian disorders. Risk alleles identified in isolated populations may, however, either be extremely rare in other populations or appear private and not observed elsewhere.<sup>17,42</sup> Findings using isolated populations may therefore not necessarily generalise to other populations, thus making replication difficult. To reduce this limitation, we selected a related British population for replication analysis, and also sought more indirect support of the findings via enrichment analysis of common variants from GWAS data and interrogation of protein–protein interaction networks within the discovery data set. Although the discovery sample was very limited in size, and

thus prone to yield spurious findings, the study identified significant associations with single variants and genes that could be replicated in the follow-up sample and/or were supported by other lines of evidence.

Among the most interesting findings, the *NOS1* missense variant is intriguing. The variant is classified by SIFT as ‘deleterious’ on transcript ENST00000338101, and while predicted as ‘benign’ by Polyphen it is classified as ‘probably damaging’ by LoFtool (Supplementary Table 2, for more details). This polymorphism shows exome-wide significance in the Faroese population and replicates in the British sample. *NOS1* encodes the neuronal nitric oxide (NO) synthase. NO is a gaseous neurotransmitter thought to have important roles in several behavioural domains. It acts as the second messenger of the *N*-methyl-D-aspartate receptor and interacts with both the dopaminergic as well as the serotonergic system.<sup>43,44</sup> Investigations of animal models and human genetic studies have implicated *NOS1* with both mood disorders and schizophrenia but with partially conflicting results.<sup>44,45</sup> Notably, the *NOS1* locus yields  $P$ -values of  $10^{-6}$  in the most recent schizophrenia GWAS.<sup>33</sup> Our findings support a role of this enzyme in bipolar disorder susceptibility and suggest the identified rare missense variant as a causal variant, providing a good basis for functional studies which relevance is further highlighted by the existence of numerous pharmaceutical targets for the mechanisms of action of NO.<sup>44,46</sup>

A synonymous variant in *FBXO21*, a gene that neighbours *NOS1*, showed similar associations. No apparent evidence from the literature seems to implicate *FBXO21* with mental disorders, and the observed association is probably due to the strong linkage disequilibrium with the *NOS1* variant ( $r^2 = 0.897$ ).

The identified *PITPNM2* missense variant was only seen in the Faroese population. *PITPNM2* encodes a phosphatidylinositol transfer protein with limited functional information. It is located in a highly significant schizophrenia GWAS locus that harbours multiple genes.<sup>33</sup> Thus, the present results not only suggest *PITPNM2* to be involved in bipolar disorder but also point to the gene as the causal culprit in this multi-gene schizophrenia locus.

Although the *P2RX7* variants fell just below the significance threshold and did not replicate in the British sample, it is worth noting that the gene has previously been associated with bipolar disorder,<sup>39,47</sup> including a study of a British sample overlapping with the present sample, showing association with another *P2RX7* variant.<sup>8</sup> The gene belongs to a family of purinoreceptors for ATP, which function as ligand-gated ion-channels, and seem to have a role in the ATP-induced glutamate transmission in the hippocampus.<sup>48</sup>

The gene-based analysis highlighted the *NCL* gene, showing exome-wide significance and replication in the British sample. *NCL* and part of its interaction network is involved in the synthesis and maturation of ribosomes.<sup>49</sup> Ribosomal DNA transcription appears to be decreased in specific cortical layers of post mortem brains in unipolar depression but not bipolar disorder.<sup>50</sup> Interestingly, this gene was also significant in our DAPPLE analysis. Furthermore, the *NCL* interaction network emphasized by the DAPPLE analysis included *PIK3C2A*, which too surpassed exome-wide significance in the Faroese population and is supported by genome-wide significance in GWAS combining bipolar disorder and schizophrenia.<sup>4</sup> Both *PIK3C2A* and *PITPNM2* are part of the phosphatidylinositol pathway, which have been widely implicated in mental disorders such as bipolar, depression and schizophrenia.<sup>4,51</sup>

Finally, the String Database<sup>52</sup> identifies an indirect interaction of *NCL*, through *MDM2* with *CREBBP*. This transcription factor is also in DAPPLE output, although not significant in terms of connectivity, and its pathway has been connected to mental illnesses in a large body of literature.<sup>53</sup> This might open opportunities to further investigate the regulation of transcription by these proteins in brain cells.

Summarizing, our results show support both for existing findings in the literature of bipolar disorder as well as for new risk genes in the disease aetiology. In particular, we identify rare variants that may provide direct leads informing on the underlying biology of bipolar disorder and schizophrenia.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

This work has been supported by the Lundbeck Foundation, through iPSYCH—the Lundbeck Foundation Initiative for Integrative Psychiatric Research. Genetic analysis of the UCL cohort has been supported by UK Medical Research Council project grants G9623693N, G0500791, G0701007 and G1000708.

## REFERENCES

- Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *Lancet* 2013; **381**: 1672–1682.
- Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012; **13**: 537–551.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.

- Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of Psychiatric Genomics Consortium et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* 2014; **19**: 1017–1024.
- Nyegaard M, Demontis D, Foldager L, Hedemand A, Flint TJ, Sørensen KM et al. CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry* 2010; **15**: 119–121.
- Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–994.
- Gratten J, Wray NR, Keller MC, Visscher PM. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci* 2014; **17**: 782–790.
- Williams HJ, Hurst JR, O'Carroll L, James C, Pao C, Chanutet E et al. The use of whole-exome sequencing to disentangle complex phenotypes. *Eur J Hum Genet* 2015; **24**: 298–301.
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- Kato T. Whole genome/exome sequencing in mood and psychotic disorders. *Psychiatry Clin Neurosci* 2015; **69**: 65–76.
- Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. *Brief Funct Genomics* 2014; **13**: 371–377.
- Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009; **41**: 35–46.
- Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. *Genome Biol* 2008; **9**: 109.
- Norio R. The Finnish Disease Heritage III: the individual diseases. *Hum Genet* 2003; **112**: 470–526.
- Peltonen L. Positional cloning of disease genes: advantages of genetic isolates. *Hum Hered* 2000; **50**: 66–75.
- Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* 2014; **512**: 190–193.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
- Jørgensen TH, Degn B, Wang AG, Vang M, Gurling H, Kalsi G et al. Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands. *Eur J Hum Genet* 2002; **10**: 381–387.
- Joensen F, Steuervald EU, Rasmussen NH. Three congenital metabolic diseases in the Faeroe Islands. Incidence, clinical and molecular genetic characteristics of Faeroese children with glycogen storage disease type IIIA, carnitine transporter deficiency and holocarboxylase synthetase deficiency. *Ugeskr Laeg* 2006; **168**: 667–670.
- Santer R, Kinner M, Steuervald U, Kjaergaard S, Skovby F, Simonsen H et al. Molecular genetic basis and prevalence of glycogen storage disease type IIIA in the Faroe Islands. *Eur J Hum Genet* 2001; **9**: 388–391.
- Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978; **35**: 837–844.
- Endicott J, Spitzer RL. Use of the research diagnostic criteria and the schedule for affective disorders and schizophrenia to study affective disorders. *Am J Psychiatry* 1979; **136**: 52–56.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–498.
- Jurinke C, van den Boom D, Cantor CR, Köster H. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* 2001; **170**: 103–116.
- Wang J, Lin M, Crenshaw A, Hutchinson A, Hicks B, Yeager M et al. High-throughput single nucleotide polymorphism genotyping using nanofluidic dynamic arrays. *BMC Genomics* 2009; **10**: 561.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012; **28**: 3326–3328.
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; **42**: 348–354.
- Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**: 311–321.

- 31 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: e1004219.
- 32 Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 2012; **28**: 1797–1799.
- 33 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
- 34 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.
- 35 Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y *et al*. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011; **7**: e1001273.
- 36 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al*. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.
- 37 Ben-David E, Shifman S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS Genet* 2012; **8**: e1002556.
- 38 Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.
- 39 McQuillin A, Bass NJ, Choudhury K, Puri V, Kosmin M, Lawrence J *et al*. Case-control studies show that a non-conservative amino-acid change from a glutamine to arginine in the P2RX7 purinergic receptor protein is associated with both bipolar- and unipolar-affective disorders. *Mol Psychiatry* 2009; **14**: 614–620.
- 40 Als TD, Jorgensen TH, Børghlum AD, Petersen PA, Mors O, Wang AG. Highly discrepant proportions of female and male Scandinavian and British Isles ancestry within the isolated population of the Faroe Islands. *Eur J Hum Genet* 2006; **14**: 497–504.
- 41 Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG *et al*. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci USA* 2011; **108**: 11983–11988.
- 42 Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J *et al*. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**: 107–116.
- 43 Weber H, Klamer D, Freudenberg F, Kittel-Schneider S, Rivero O, Scholz C-J *et al*. The genetic contribution of the NO system at the glutamatergic post-synapse to schizophrenia: further evidence and meta-analysis. *Eur Neuropsychopharmacol* 2014; **24**: 65–85.
- 44 Freudenberg F, Altoa A, Reif A. Neuronal nitric oxide synthase (NOS1) and its adaptor, NOS1AP, as a genetic risk factors for psychiatric disorders. *Genes Brain Behav* 2015; **14**: 46–63.
- 45 Buttenschön HN, Mors O, Ewald H, McQuillin A, Kalsi G, Lawrence J *et al*. No association between a neuronal nitric oxide synthase (NOS1) gene polymorphism on chromosome 12q24 and bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2004; **124B**: 73–75.
- 46 Doucet MV, Harkin A, Dev KK. The PSD-95/nNOS complex: new drugs for depression? *Pharmacol Ther* 2012; **133**: 218–229.
- 47 Barden N, Harvey M, Gagné B, Shink E, Tremblay M, Raymond C *et al*. Analysis of single nucleotide polymorphisms in genes in the chromosome 12Q24.31 region points to P2RX7 as a susceptibility gene to bipolar affective disorder. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B**: 374–382.
- 48 Andó RD, Sperlágh B. The role of glutamate release mediated by extrasynaptic P2X7 receptors in animal models of neuropathic pain. *Brain Res Bull* 2013; **93**: 80–85.
- 49 Roger B, Moisan D, Amalric F, Bouvet P. Nucleolin provides a link between RNA polymerase I transcription and pre-ribosome assembly. *Chromosoma* 2003; **111**: 399–407.
- 50 Gos T, Steiner J, Krell D, Bielau H, Mawrin C, Krzyżanowski M *et al*. Ribosomal DNA transcription in the anterior cingulate cortex is decreased in unipolar but not bipolar I depression. *Psychiatry Res* 2013; **210**: 338–345.
- 51 Lang UE, Puls I, Müller DJ, Strutz-Seeböhm N, Gallinat J. Molecular mechanisms of schizophrenia. *Cell Physiol Biochem* 2007; **20**: 687–702.
- 52 Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Müller J *et al*. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009; **37**: D412–D416.
- 53 Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015; **18**: 199–209.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Supplementary Information accompanies the paper on the *Translational Psychiatry* website (<http://www.nature.com/tp>)